

CLAIMS

What is claimed is:

1. A page index system comprising:
a page data store that stores reference information associated with pages; and,
a crawler component that receives a page, retrieves reference information associated with the page from the page data store, and, provides the page and associated reference information.
2. A web crawler employing the system of claim 1.
3. The system of claim 1, the reference information comprising anchor text.
4. The system of claim 3, the reference information comprising at least one of a sentence fragment, a sentence and a paragraph in proximity to a referencing uniform resource locator.
5. An Internet search engine employing the page and associated reference information provided by the system of claim 1.
6. The system of claim 1, the page data store storing a uniform resource locator identifying a page, the uniform resource locator further being employed to store the reference information associated with a particular page.
7. One or more readable media having stored thereon computer executable instructions for carrying out the system of claim 1.
8. A crawler comprising:
an input component that receives a page;
a parser component that parses the page for another page referenced on the page, stores reference information associated with the another page in a page data store;

a retrieval component that receives the page and retrieves reference information associated with the page from the page data store; and,

an output component that provides an output comprising the page merged with the reference information associated with the page.

9. A page indexing system comprising the crawler of claim 8.
10. The page indexing system of claim 9 further comprising the page data store.
11. The page indexing system of claim 10, the page data store storing reference information associated with pages.
12. The system of claim 11, the page data store storing a uniform resource locator identifying a page, the uniform resource locator further being employed to store the reference information associated with a particular page.
13. The crawler of claim 8, the reference information comprising anchor text.
14. The crawler of claim 13, the reference information comprising at least one of a sentence fragment, a sentence and a paragraph in proximity to a referencing uniform resource locator.
15. A method facilitating page indexing comprising:
retrieving reference information associated with a page; and,
providing an output comprising the page merged with the reference information associated with the page.
16. The method of claim 15, further comprising at least one of the following:
receiving a request to retrieve the page;
retrieving the page; and,

storing reference information associated with a uniform resource locator on a page.

17. The method of claim 15, retrieval of the reference information associated with the page being based, at least in part, upon a uniform resource locator identifying the page.

18. One or more computer readable media having stored thereon computer executable instructions for carrying out the method of claim 15.

19. A memory for storing data for access by an application program being executed on a page indexing system, comprising:

a data structure stored in said memory, said data structure a first data field comprising reference information associated with a page; and,
a second data field comprising the page.

20. The memory of claim 19, the reference information comprising anchor text.

21. The memory of claim 20, the reference information further comprising at least one of a sentence fragment, a sentence and a paragraph in proximity to a referencing uniform resource locator.

22. One or more computer readable media storing computer executable components of a crawler comprising:

an input component that receives a page;
a parser component that parses the page for another page referenced on the page, stores reference information associated with the another page in a page data store;
a retrieval component that receives the page and retrieves reference information associated with the page from the page data store; and,
an output component that provides an output comprising the page merged with the reference information associated with the page.

23. The media of claim 22, the page data store storing a uniform resource locator identifying a page, the uniform resource locator further being employed to store the reference information associated with a particular page.
24. The media of claim 22, the reference information comprising anchor text.
25. The media of claim 22, the reference information comprising at least one of a sentence fragment, a sentence and a paragraph in proximity to a referencing uniform resource locator.
26. A page index system comprising:
means for storing reference information associated with pages; and,
means for receiving a page;
means for retrieving reference information associated with the page from means for storing reference information; and,
means for providing an output, the output comprising the page merged with the reference information associated with the page.
27. The system of claim 26, the means for storing reference information storing a uniform resource locator identifying a page, the uniform resource locator further being employed to store the reference information associated with a particular page.
28. The media of claim 26, the reference information comprising anchor text.
29. The media of claim 26, the reference information comprising at least one of a sentence fragment, a sentence and a paragraph in proximity to a referencing uniform resource locator.